

Scale checks in censored regression

Holger Dette

*Fakultät für Mathematik
Ruhr-Universität Bochum*

Cédric Heuchenne

*HEC-Ecole de gestion de l'ULg
Université de Liège*

October 9, 2009

Abstract

Suppose the random vector (X, Y) satisfies the regression model $Y = m(X) + \sigma(X)\varepsilon$, where $m(\cdot)$ and $\sigma(\cdot)$ are unknown location and scale functions and ε is independent of X . The response Y is subject to random right censoring and the covariate X is completely observed. A new test for a specific parametric form of any scale function $\sigma(\cdot)$ (including the standard deviation function) is proposed. Its statistic is based on the distribution of the residuals obtained from the assumed regression model. Weak convergence of the corresponding process is obtained and its finite sample behaviour is studied via simulations. Finally, characteristics of the test are illustrated in the analysis of a fatigue data set.

KEY WORDS: Bootstrap; Goodness-of-fit tests; Fatigue data; Kernel method; Least squares estimation; Nonparametric regression; Right censoring; Survival analysis.

1 Introduction

In this paper, we consider the following heteroscedastic regression model

$$Y = m(X) + \sigma(X)\varepsilon, \quad (1.1)$$

where $m(X)$ and $\sigma(X)$ are some unknown but smooth location and scale functions and ε (with location zero and scale one) is independent of X (one-dimensional). Suppose also that Y is subject to random right censoring, i.e. instead of observing Y , we only observe (Z, Δ) , where $Z = \min(Y, C)$, $\Delta = I(Y \leq C)$ and the random variable C represents the censoring time which is independent of Y , conditionally on X . Let $(Y_i, C_i, X_i, Z_i, \Delta_i)$ ($i = 1, \dots, n$) be n independent copies of (Y, C, X, Z, Δ) .

The aim of this paper is to test the hypothesis

$$H_0 : \sigma(\cdot) \in \mathcal{M} \text{ versus } H_1 : \sigma(\cdot) \notin \mathcal{M}, \quad (1.2)$$

where $\mathcal{M} = \{\sigma_\vartheta : \vartheta \in \Theta\}$ is a class of parametric functions and $\Theta \subset \mathbb{R}^D$. However, it is well known that functions which involve the right tails of the conditional distribution of $F(\cdot|x) = P(Y \leq \cdot | X = x)$ of Y given $X = x$ (like the conditional variance $Var[Y|X = x] = \int (y - E[Y|x])^2 dF(y|x)$) cannot be estimated in a consistent way in a completely nonparametric model, due to the presence of right censoring. In fact, the completely nonparametric (kernel) estimator of $F(\cdot|x)$ is not consistent in the right tail (see Beran, 1981) if the conditional distribution of Y has a larger support than the conditional distribution of C . In this paper, we present a way to overcome this problem by imposing the weak model assumption (1.1) and replacing the class \mathcal{M} in (1.2) by the more specific class

$$\mathcal{M}' = \{\sigma_\vartheta : \sigma_\vartheta = \vartheta_1 \sigma_{1p} \text{ and } \vartheta \in \Theta\}$$

(H_0 will therefore be replaced by a new hypothesis H'_0 using \mathcal{M}'), where $p = (\vartheta_2 \dots \vartheta_D)'$, $\vartheta = (\vartheta_1 p)'$ and $\Theta \subset \mathbb{R}^D$. We will show that using those assumptions enables to reduce the inconsistency problems included in the testing procedure.

A nice advantage of the method is that it applies to any scale function (see Section 2 for formal definitions of location and scale functions) including the square root of the general expression

$$\sigma^2(x) = a_0^\sigma \int_0^1 (F^{-1}(s|x) - m_0(x))^2 L^\sigma(s) ds + \sum_{j=1}^{k^\sigma} a_j^\sigma \left\{ \int_0^1 \rho_j(F^{-1}(s|x) - m_j(x)) ds \right\}^2, \quad (1.3)$$

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the quantile function of Y given x , m_0, \dots, m_{k^σ} are general location functions of the type

$$m_p(x) = a_{p0}^m \int_0^1 F^{-1}(s|x) L_p^m(s) ds + \sum_{j=1}^{k_p^m} a_{pj}^m F^{-1}(s_{pj}^m|x), \quad (1.4)$$

$p = 0, \dots, k^\sigma$ (see e.g. Serfling, 1980, p. 265), $L^\sigma(s)$ and $L_p^m(s)$ are given weight functions satisfying $\int_0^1 L^\sigma(s)ds = 1$ and $\int_0^1 L_p^m(s)ds = 1$, $p = 0, \dots, k^\sigma$, $k^\sigma \geq 0$, $k_p^m \geq 0$, $a_0^\sigma, \dots, a_{k^\sigma}^\sigma$ are positive real numbers (a_0^σ can be zero if $k^\sigma > 0$), $a_{p0}^m, \dots, a_{pk_p^m}^m$ are real numbers such that $\sum_{j=0}^{k_p^m} a_{pj}^m = 1$, $p = 0, \dots, k^\sigma$, $\rho_j(u) = s_j^\sigma uI(u \geq 0) + (s_j^\sigma - 1)uI(u < 0)$, $j = 1, \dots, k^\sigma$, and $0 < s_1^\sigma, \dots, s_{k^\sigma}^\sigma, s_{p1}^m, \dots, s_{pk_p^m}^m < 1$, $p = 0, \dots, k^\sigma$. The expression (1.3) includes a very broad class of scale functions. For example, for $k^\sigma = 0$, $k_0^m = 0$, $L^\sigma \equiv L_0^m \equiv 1$, (1.3) corresponds to the conditional variance, for $a_0^\sigma = a_{10}^m = 0$, $a_1^\sigma = k^\sigma = k_1^m = 1$, $s_1^\sigma = 1/2$ ($s_1^\sigma = 1/4$) and $s_{11}^m = 1/2$ ($s_{11}^m = 1/4$), (1.3) corresponds to the square of the usual scale function associated with the median (the first quartile).

A key idea to obtain consistent test statistics is to replace (possibly inconsistently estimated) $m^P(x) = (m_0(x) \cdots m_{k^\sigma}(x))'$ and $\sigma(x)$ in (1.4) and (1.3) by other location and scale functions $m^0(x)$ and $\sigma^0(x)$ of the same type as (1.4) and (1.3) and which can be consistently estimated (see Section 2 for specific definitions). Indeed, under model (1.1), it is easy to check that $\sigma^0(X)$ and $\sigma(X)$ are proportional and that the model

$$Y = m^0(X) + \sigma^0(X)\varepsilon^0 \quad (1.5)$$

also holds for any location and scale functions $m^0(X)$ and $\sigma^0(X)$ with $\varepsilon^0 = (Y - m^0(X))/\sigma^0(X)$ independent of X .

The approach used in this paper is based on the estimated difference of residuals distributions given by

$$\hat{F}_\varepsilon^0(y) - \hat{F}_{\varepsilon_0}(y), \quad -\infty < y \leq T, \quad (1.6)$$

where $\hat{F}_\varepsilon^0(\cdot)$ and $\hat{F}_{\varepsilon_0}(\cdot)$ are estimators (described in Section 2) of $F_\varepsilon^0(y) = P(\varepsilon^0 \leq y)$ and $F_{\varepsilon_0}(y) = P(\varepsilon_0^0 \leq y)$, the distributions of the residuals obtained from model (1.5),

$$\varepsilon^0 = \frac{Y - m^0(X)}{\sigma^0(X)} \quad (1.7)$$

and

$$\varepsilon_0^0 = \frac{Y - m^0(X)}{\sigma_{\hat{\theta}_0}(X)}. \quad (1.8)$$

$\sigma_{\hat{\theta}_0}(\cdot)$ denotes the best approximation of $\sigma^0(\cdot)$ by elements of the class \mathcal{M}' (if H'_0 is true, then $\sigma_{\hat{\theta}_0}(X) = \sigma^0(X) = \sigma_{\theta_0}(X)$, for a true parametric scale $\sigma_{\theta_0}(X)$). The point T in (1.6) is chosen smaller than the upper bound of the support of the distribution of the observed residuals $\frac{Z - m^0(X)}{\sigma^0(X)}$ (see Section 2). The presence of this cutting point is due to possible right censoring problems of the residuals (1.7) and (1.8). However, under model (1.5), each false H'_0 will be detected by a nonzero difference between both residuals distributions $F_\varepsilon^0(y)$ and $F_{\varepsilon_0}(y)$ for points y smaller than T (see Lemma 2.1).

In the case of no censoring the problem of testing for heteroscedasticity in the classical nonparametric regression model with conditional expectation m and conditional variance σ^2 has been considered by numerous authors [see Dette and Munk (1998), Dette (2002), Liero (2003), Dette and Hetzler (2009a,b) and Dette, Neumeyer and Van Keilegom (2007) among others]. Similar testing problems in semiparametric models have been considered by You and Chen (2005).

Although a number of goodness-of-fit tests exists for the regression function with censored data, few results are obtained for the conditional variance and especially for a scale function which is different from the usual standard deviation function. In the censored case, González Manteiga, Heuchenne and Sánchez Sellero (2007) considered goodness-of-fit tests for the conditional mean and variance functions while Pardo Fernández, Van Keilegom and González Manteiga (2007) addressed the problem for a specific location function using the process of the difference of residuals distributions. This process has been widely studied, b.e., by Dette, Pardo Fernández and Van Keilegom (2007) or Van Keilegom, González Manteiga and Sánchez Sellero (2007). Indeed, it is more naturally related to the commonly used graphical procedures based on visual examination of the residuals (see Atkinson 1985). In the case of variance testing, it has been used by Dette, Neumeyer and Van Keilegom (2007). In fact, nonparametric residuals submitted to compressions and expansions (due to their transformation into parametric residuals) seem to produce important discrepancies in the corresponding distributions.

The paper is organized as follows. In the next section, the testing procedure is described in detail. Section 3 summarizes the main asymptotic results, including the weak convergence (under H'_0) of the proposed process to a Gaussian process. In Section 4, we present the results of a simulation study and different parametric forms are tested on a fatigue data set in Section 5. Finally, the Appendix contains the assumptions, functions and proofs needed to obtain the main results of Section 3.

Remark 1.1 (Choice of the null hypothesis). In practice, a null hypothesis for which a parametric form σ_{ϑ} cannot be factorized by one of its parameters can be tested. In this case, the above methodology doesn't apply. However, this restriction is not so much constraining since a new null hypothesis for which σ_{ϑ} is multiplied by a parameter can be tested; rejection of this will also imply rejection of the initial null hypothesis.

2 Notations and description of the method

As explained in Section 1, the idea of the method is first to construct consistent residuals by replacing $m(\cdot)$ and $\sigma(\cdot)$ by some specific $m^0(\cdot)$ and $\sigma^0(\cdot)$ and then to introduce the

so-obtained residuals in expression (1.6). To develop the procedure, we first need to introduce a number of notations.

Define $F(y|x) = P(Y \leq y|x)$, $G(y|x) = P(C \leq y|x)$, $H(y|x) = P(Z \leq y|x)$, $H(y) = P(Z \leq y)$, $H_\delta(y|x) = P(Z \leq y, \Delta = \delta|x)$ ($\delta = 0, 1$), $F_X(x) = P(X \leq x)$, $S_\varepsilon^0(y) = 1 - F_\varepsilon^0(y)$, $S_{\varepsilon_0}(y) = 1 - F_{\varepsilon_0}(y)$, for $E^0 = (Z - m^0(X))/\sigma^0(X)$, we denote $H_\varepsilon^0(y) = P(E^0 \leq y)$, $H_{\varepsilon\delta}^0(y) = P(E^0 \leq y, \Delta = \delta)$, $H_\varepsilon^0(y|x) = P(E^0 \leq y|x)$, $H_{\varepsilon\delta}^0(y|x) = P(E^0 \leq y, \Delta = \delta|x)$, for $E_0^0 = (Z - m^0(X))/\sigma_{\theta_0}^0(X)$, we denote $H_{\varepsilon_0}(y) = P(E_0^0 \leq y)$, $H_{\varepsilon_0\delta}(y) = P(E_0^0 \leq y, \Delta = \delta)$, $H_{\varepsilon_0}(y|x) = P(E_0^0 \leq y|x)$, $H_{\varepsilon_0\delta}(y|x) = P(E_0^0 \leq y, \Delta = \delta|x)$, and for $C^0 = (C - m^0(X))/\sigma^0(X)$ (resp. $C_0^0 = (C - m^0(X))/\sigma_{\theta_0}^0(X)$), we denote $G_\varepsilon^0(y) = P(C^0 \leq y)$ (resp. $G_{\varepsilon_0}(y) = P(C_0^0 \leq y)$). The probability density functions of the distributions defined above will be denoted with lower case letters and R_X denotes the compact support of the distribution of the random variable X .

Now, let $m^l(\cdot)$ be any location function and $\sigma^s(\cdot)$ be any scale function, meaning that $m^l(x) = T(F(\cdot|x))$ and $\sigma^s(x) = S(F(\cdot|x))$ for some functionals T and S that satisfy $T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b$ and $S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x))$, for all $a \geq 0$ and $b \in \mathbb{R}$ (here $F_{aY+b}(\cdot|x)$ denotes the conditional distribution of $aY + b$ given $X = x$). Let $\varepsilon^{ls} = (Y - m^l(X))/\sigma^s(X)$. Then, it can be easily seen that if model (1.1) holds (i.e. ε is independent of X), then ε^{ls} is also independent of X . Moreover, $\sigma(X) = S^d(F_{\varepsilon^{ls}}(\cdot))\sigma^s(X)$, $F_{\varepsilon^{ls}}(\cdot)$ denoting the distribution of ε^{ls} and $S^d(F_{\varepsilon^{ls}}(\cdot))$ the scale functional corresponding to $\sigma(X)$ and applied to $F_{\varepsilon^{ls}}(\cdot)$ instead of $F(\cdot|X)$ (ε^{ls} independent of X). Therefore, achieving goodness-of-fit tests for $\sigma(\cdot)$ or $\sigma^s(\cdot)$ are equivalent when the parametric function to test can be factorized by one of its parameters. If the objective is to estimate $\sigma(X)$, this can be achieved in a second step by estimating the quantity $S^d(F_{\varepsilon^{ls}}(\cdot))$ (a simple case for $S^d(F_{\varepsilon^{ls}}(\cdot))$ is the standard deviation of the residuals).

Next, for $m^l(x)$ and $\sigma^{s2}(x)$, we choose

$$m^0(x) = \int_0^1 F^{-1}(s|x)J(s) ds, \quad \sigma^{02}(x) = \int_0^1 F^{-1}(s|x)^2 J(s) ds - m^{02}(x), \quad (2.1)$$

where $J(s)$ is a given score function satisfying $\int_0^1 J(s) ds = 1$. When $J(s)$ is chosen appropriately (namely put to zero in the right tail, there where the quantile function cannot be estimated in a consistent way due to the right censoring), $m^0(x)$ and $\sigma^0(x)$ can be estimated consistently (see Section 4 for a data-driven choice of J). The distribution $F(y|x)$ in (2.1) is replaced by the Beran (1981) estimator, defined by (in the case of no ties) :

$$\hat{F}(y|x) = 1 - \prod_{Z_i \leq y, \Delta_i = 1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, a_n)} \right\}, \quad (2.2)$$

where

$$W_i(x, a_n) = \frac{K\left(\frac{x-X_i}{a_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{a_n}\right)},$$

K is a kernel function and $\{a_n\}$ a bandwidth sequence. Therefore,

$$\hat{m}^0(x) = \int_0^1 \hat{F}^{-1}(s|x)J(s) ds, \quad \hat{\sigma}^{02}(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s) ds - \hat{m}^{02}(x) \quad (2.3)$$

estimate $m^0(x)$ and $\sigma^{02}(x)$. Next,

$$\hat{F}_\varepsilon^0(y) = 1 - \prod_{\hat{E}_{(i)}^0 \leq y, \Delta_{(i)}=1} \left(1 - \frac{1}{n-i+1}\right), \quad (2.4)$$

denotes the Kaplan-Meier (1958)-type estimator of F_ε^0 (in the case of no ties), where $\hat{E}_i^0 = (Z_i - \hat{m}^0(X_i))/\hat{\sigma}^0(X_i)$, $\hat{E}_{(i)}^0$ is the i -th order statistic of $\hat{E}_1^0, \dots, \hat{E}_n^0$ and $\Delta_{(i)}$ is the corresponding censoring indicator. This estimator has been studied in detail by Van Keilegom and Akritas (1999).

Next, we consider a parametric estimator for σ^0 defined by

$$\vartheta_n := \operatorname{argmin}_{\vartheta \in \Theta} \sum_{i=1}^n [\hat{\sigma}^0(X_i) - \sigma_\vartheta(X_i)]^2. \quad (2.5)$$

Asymptotic properties of this estimator can be found in the Appendix. Similarly to (2.4), let

$$\hat{F}_{\varepsilon_0}(y) = 1 - \prod_{\hat{E}_{(i)0}^0 \leq y, \Delta_{(i)0}=1} \left(1 - \frac{1}{n-i+1}\right), \quad (2.6)$$

denote the Kaplan-Meier (1958)-type estimator of F_{ε_0} (in the case of no ties), where $\hat{E}_{i,0}^0 = (Z_i - \hat{m}^0(X_i))/\sigma_{\vartheta_n}(X_i)$, $\hat{E}_{(i)0}^0$ is the i -th order statistic of $\hat{E}_{1,0}^0, \dots, \hat{E}_{n,0}^0$ and $\Delta_{(i)0}$ is the corresponding censoring indicator.

Therefore, we consider the following process

$$\hat{W}(y) = n^{1/2}(\hat{F}_\varepsilon^0(y) - \hat{F}_{\varepsilon_0}(y)), \quad -\infty < y \leq T, \quad (2.7)$$

where $T < \tau_{H_\varepsilon^0} = \tau_{F_\varepsilon^0} \wedge \tau_{G_\varepsilon^0}$ and $\tau_F = \inf\{x : F(x) = 1\}$. As it is clear from the definitions of $\hat{F}_\varepsilon^0(y)$ and $\hat{F}_{\varepsilon_0}(y)$, $\hat{W}(y)$ is actually estimating

$$W(y) = n^{1/2}(F_\varepsilon^0(y) - F_{\varepsilon_0}(y)) \quad (2.8)$$

for $-\infty < y \leq T$, such that the whole supports of the involved distributions are not considered. However, as already mentioned in Section 1, differences between scale functions can only be detected with a part of the considered supports. This is shown by the following Lemma.

Lemma 2.1 *Assume that all moments of the random variable $\varepsilon^0 I(\varepsilon^0 \leq T)$ exist and that T is a positive real value. Then H'_0 holds if and only if there exists some $\tilde{\theta}_0 \in \Theta$ such that $F_\varepsilon^0(y \wedge T) = F_{\varepsilon_0}(y \wedge T)$ for all y , $-\infty < y < \infty$.*

Proof of Lemma 2.1. The direct implication is trivial. On the other hand, assume that there exists some $\tilde{\theta}_0$ such that $F_\varepsilon^0(y \wedge T) = F_{\varepsilon_0}(y \wedge T)$, for all y and some T . It is then clear that

$$P(\varepsilon^0 \leq y) = P(\varepsilon_0^0 \leq y) = P(\varepsilon^0 \leq y \frac{\sigma_{\tilde{\theta}_0}(X)}{\sigma^0(X)}) \text{ for all } y \leq T. \quad (2.9)$$

We have

$$\begin{aligned} \int_{-\infty}^T y^{2n} dF_\varepsilon^0(y) &= \int_{-\infty}^T y^{2n} dF_{\varepsilon_0}(y) \\ &= \int \int_{-\infty}^{T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)}} \left(y \frac{\sigma^0(x)}{\sigma_{\tilde{\theta}_0}(x)} \right)^{2n} dF_\varepsilon^0(y) dF_X(x) \\ &\leq \int \left(\frac{\sigma^{02}(x)}{\sigma_{\tilde{\theta}_0}^2(x)} \right)^n \int_{-\infty}^T y^{2n} dF_\varepsilon^0(y) dF_X(x), \end{aligned} \quad (2.10)$$

for all $n \in \mathbb{N}$, since $\int_T^{T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)}} \left(y \frac{\sigma^0(x)}{\sigma_{\tilde{\theta}_0}(x)} \right)^{2n} dF_\varepsilon^0(y) dF_X(x) \leq 0$. The last inequality can be obtained as follows. For regions of R_X where $\frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)} < 1$,

$$\int_T^{T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)}} \left(y \frac{\sigma^0(x)}{\sigma_{\tilde{\theta}_0}(x)} \right)^{2n} dF_\varepsilon^0(y) \leq - \int_T^{T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)}} T^{2n} dF_\varepsilon^0(y),$$

while for regions of R_X where $\frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)} \geq 1$,

$$\int_T^{T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)}} \left(y \frac{\sigma^0(x)}{\sigma_{\tilde{\theta}_0}(x)} \right)^{2n} dF_\varepsilon^0(y) \leq \int_T^{T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)}} T^{2n} dF_\varepsilon^0(y).$$

Using (2.9), we therefore have

$$\int_{R_X} F_\varepsilon^0 \left(T \frac{\sigma_{\tilde{\theta}_0}(x)}{\sigma^0(x)} \right) dF_X(x) - F_\varepsilon^0(T) = 0.$$

The inequality $1 \geq \int \left(\frac{\sigma^{02}(x)}{\sigma_{\tilde{\theta}_0}^2(x)} \right)^n dF_X(x)$ is simply obtained by replacing T by $-T$ in (2.10). Therefore we obtain from Carleman's condition (see e.g. Feller (1966) p. 228) that the distribution of the random variable $\frac{\sigma^{02}(X)}{\sigma_{\tilde{\theta}_0}^2(X)}$ coincides with the distribution of the constant random variable $U \equiv 1$, that is H'_0 holds.

From (2.7), we therefore propose a Kolmogorov-Smirnov type statistic

$$T_{KS} = \sup_{-\infty < y \leq T} |\hat{W}(y)|$$

and a Cramer-von Mises type statistic

$$T_{CM} = \frac{1}{\hat{F}_{\varepsilon}^0(T)} \int_{-\infty}^T \hat{W}^2(y) d\hat{F}_{\varepsilon}^0(y).$$

The null hypothesis H'_0 is rejected for large values of the test statistics.

Remark 2.2. This testing procedure is used to check a parametric form for a scale function $\sigma(\cdot)$ but (2.5) only provides a parametric estimation for $\sigma^0(\cdot)$. However, a parametric estimation for $\sigma(\cdot)$ under H'_0 is easily obtained by multiplying $\sigma_{\vartheta_n}(\cdot)$ by an estimator of $S^d(F_{\varepsilon_0}(\cdot))$ which in the case of (1.3), (1.4), could be given by

$$\begin{aligned} \tilde{S}^{d2}(\hat{F}_{\varepsilon_0}(\cdot)) &= a_0^\sigma \int_{-\infty}^T (y - \tilde{T}_0(\hat{F}_{\varepsilon_0}(\cdot)))^2 L^\sigma(\hat{F}_{\varepsilon_0}(y)) d\hat{F}_{\varepsilon_0}(y) \\ &\quad + \sum_{j=1}^{k^\sigma} a_j^\sigma \left\{ \int_0^1 \rho_j(\hat{F}_{\varepsilon_0}^{-1}(s) \wedge T - \tilde{T}_j(\hat{F}_{\varepsilon_0}(\cdot))) ds \right\}^2, \end{aligned} \quad (2.11)$$

where $\tilde{T}_0, \dots, \tilde{T}_{k^\sigma}$ are pseudo-location functionals of the type

$$\tilde{T}_q(\hat{F}_{\varepsilon_0}(\cdot)) = a_{q0}^m \int_{-\infty}^T y L_q^m(\hat{F}_{\varepsilon_0}(y)) d\hat{F}_{\varepsilon_0}(y) + \sum_{j=1}^{k_q^m} a_{qj}^m (\hat{F}_{\varepsilon_0}^{-1}(s_{qj}^m) \wedge T), \quad (2.12)$$

$q = 0, \dots, k^\sigma$.

3 Asymptotic results

We start by developing an asymptotic representation for the expression (2.7) under the null hypothesis H'_0 and where the remaining term is $o_P(n^{-1/2})$ uniformly in y . This will allow us to obtain the weak convergence of the process $\hat{W}(y)$. Finally, the asymptotic distributions of the proposed test statistics are obtained under the null hypothesis H'_0 . The assumptions, proofs and involved functions in the results below are given in the Appendix.

Theorem 3.1 *Assume (A1)-(A8) (in the Appendix). Then, under the null hypothesis H'_0 ,*

$$\hat{F}_{\varepsilon}^0(y) - \hat{F}_{\varepsilon_0}(y) = n^{-1} \sum_{i=1}^n \chi_{\theta_0}(X_i, Z_i, \Delta_i, y) + R_n(y),$$

where $\sup\{|R_n(y)|; -\infty < y \leq T\} = o_P(n^{-1/2})$ and $\chi_{\theta_0}(x, z, \delta, y)$ is defined in the Appendix.

Theorem 3.2 Assume (A1)-(A8) (in the Appendix). Then, under the null hypothesis H'_0 , the process $\hat{W}(y) = n^{-1/2}(\hat{F}_\varepsilon^0(y) - \hat{F}_{\varepsilon_0}(y))$, $-\infty < y \leq T$ converges weakly to a centered gaussian process $W(y)$ with covariance function

$$\text{Cov}(W(y), W(y')) = E[\chi_{\theta_0}(X, Z, \Delta, y)\chi_{\theta_0}(X, Z, \Delta, y')].$$

Corollary 3.3 Assume (A1)-(A8) (in the Appendix). Then, under the null hypothesis H'_0 ,

$$T_{KS} \xrightarrow{d} \sup_{-\infty < y \leq T} |W(y)|,$$

$$T_{CM} \xrightarrow{d} \frac{1}{F_\varepsilon^0(T)} \int_{-\infty}^T W^2(y) dF_\varepsilon^0(y).$$

4 Practical implementation and simulations

In this section, we study the finite sample behavior of both test statistics. We are interested in the behavior of the percentage of simulated samples for which the null hypothesis is rejected. The simulations are carried out for samples of size $n = 50$ and $n = 100$ and the results are obtained by using 1000 simulation runs. First, we describe the characteristics of the proposed methods.

- (1) For the score function J , we recommend the choice

$$J(s) = b^{-1}I(0 \leq s \leq b) \quad (0 \leq s \leq 1),$$

where

$$b = \min_{1 \leq i \leq n} \hat{F}(+\infty | X_i).$$

In this way, the region where the Beran estimator is inconsistent is not used, and on the other hand, we exploit to a maximum the common ‘consistent’ region.

- (2) For the $K(x)$, we work with the biquadratic kernel function

$$K(x) = (15/16)(1 - x^2)^2 I(|x| \leq 1).$$

In order to improve the behavior near the boundaries of the covariate space, we use the reflection method to compute all kernel estimates.

- (3) For the calculation of the parametric estimate in (2.5) we use the Levenberg-Marquardt algorithm (Levenberg (1944) and Marquardt (1963)) (for a fixed value of the bandwidth parameter).

- (3) The point T is chosen as $\hat{E}_{(n)0}^0$ for the Cramer von Mises test and as $\hat{E}_{(n)}^0 \vee \hat{E}_{(n)0}^0$ for the Kolmogorov-Smirnov type statistic.

For the calculation of the critical values we need the distributions of the statistics T_{KS} and T_{CM} under the null hypothesis. Unfortunately, the asymptotic distributions obtained in Corollary 3.3 are too complicated and contain too many unknown quantities. We therefore propose a bootstrap procedure to estimate the critical values of the test in practical situations. This is based on a smoothed version of the 'naive bootstrap' described in Efron (1981) and on the method suggested in Pardo Fernández, Van Keilegom and González Manteiga (2007).

First, define $\tilde{E}_1^0, \dots, \tilde{E}_n^0$, the standardized versions of the residuals $\hat{E}_1^0, \dots, \hat{E}_n^0$. In fact, for $\lambda_1 = \int y J(\hat{F}_\varepsilon^0(y)) d\hat{F}_\varepsilon^0(y)$ and $\lambda_2^2 = \int (y - \lambda_1)^2 J(\hat{F}_\varepsilon^0(y)) d\hat{F}_\varepsilon^0(y)$, we compute $\tilde{E}_i^0 = (\hat{E}_i^0 - \lambda_1)/\lambda_2$, $i = 1, \dots, n$. Note that the right tail of $\hat{F}_\varepsilon^0(y)$ could be involved by λ_1 and λ_2 . In this case, the last order statistic $\hat{E}_{(n)}^0$ is redefined as uncensored. The bootstrap procedure consists of the following steps. For fixed B and $b = 1, \dots, B$,

- (1) For $i = 1, \dots, n$:

· Let

$$Y_{i,b}^* = \hat{m}^0(X_i) + \sigma_{\vartheta_n}(X_i) \varepsilon_{i,b}^*,$$

where $\varepsilon_{i,b}^* = V_{i,b} + aS_{i,b}$, $V_{i,b}$ is drawn from \tilde{F}_ε^0 , (the Kaplan-Meier estimator based on the standardized residuals) and $S_{i,b}$ is a normal distributed random variable with mean 0 and variance 1 which introduces a small perturbation in the residuals (controlled by the constant a).

- Select $C_{i,b}^*$ from a smoothed version of $\hat{G}(\cdot|X_i)$, the Beran (1981) estimator of the distribution $G(\cdot|X_i)$ obtained by replacing Δ_i by $1 - \Delta_i$ in the expression of $\hat{F}(\cdot|X_i)$.
 · Let $Z_{i,b}^* = \min(Y_{i,b}^*, C_{i,b}^*)$ and $\Delta_{i,b}^* = I(Y_{i,b}^* \leq C_{i,b}^*)$.

- (2) The bootstrap sample is $\{(X_i, Z_{i,b}^*, \Delta_{i,b}^*), i = 1, \dots, n\}$.

- (3) Let $T_{KS,b}^*$ and $T_{CM,b}^*$ be the test statistics calculated with the bootstrap sample.

Let $T_{KS,(b)}^*$ be the b -th order statistic of $T_{KS,1}^*, \dots, T_{KS,B}^*$, and analogously for $T_{CM,(b)}^*$. Then $T_{KS,((1-\alpha)B+1)}^*$ and $T_{CM,((1-\alpha)B+1)}^*$ (where $[\cdot]$ denotes the integer part) approximate the $(1 - \alpha)$ -quantiles of the distributions of T_{KS} and T_{CM} . In the following discussion we illustrate the finite sample properties of this procedure for the Cramer von Mises statistic in two examples. The results for the Kolmogoroff Smirnov case are similar and therefore omitted.

In our first example we consider the problem of testing for homoscedasticity in the regression model

$$Y_i = X_i + 0.5 \exp(cX_i) \xi_i, \quad i = 1, \dots, n,$$

where X_1, \dots, X_n are i.i.d. $\sim \mathcal{U}[0, 1]$ and ξ_1, \dots, ξ_n are i.i.d. $\sim \mathcal{N}[0, 1]$. Note that the case $c = 0$ corresponds to the null hypothesis of homoscedasticity and two alternatives are considered, i.e. $c = 0.5$ and $c = 1$. The censoring times are generated by the model

$$C_i = X_i + q + 0.5 \exp(cX_i) \eta_i, \quad i = 1, \dots, n,$$

where q determines the amount of censoring and η_1, \dots, η_n are again $\sim \mathcal{N}[0, 1]$.

In Table 1, we display the simulated rejection probabilities based on 1000 simulation runs, where the number of bootstrap replications is chosen as $B = 199$. For the bandwidth in the conditional Kaplan Meier estimate we used $a_n = 0.1$, while the bandwidth $a_n = 0.15$ was used in the Beran estimate to generate the censored observations in the bootstrap. For the constant q we considered the cases $q = 9^9, q = 0.85$ and $q = 0.35$ which corresponds to an amount of 0%, 11% and 31% censoring under the null hypothesis, respectively. Under the alternative $c = 0.5$ and $c = 1$ the cases $q = 9^9, q = 0.85, q = 0.35$ yield to 0%, 18% and 35% and 0%, 24% and 39% censoring, respectively.

Table 1 *Simulated rejection probabilities of the bootstrap test for the hypothesis of homoscedasticity.*

n		50				100			
c	q	2.5%	5%	10%	20%	2.5%	5%	10%	20%
0	9^9	4.8	6.2	11.5	22.1	3.7	5.8	10.1	21.4
	0.85	4.6	6.6	10.9	18.2	2.1	4.2	8.5	16.3
	0.35	4.7	6.9	11.8	17.5	3.9	6.2	10.7	19.4
0.5	9^9	19.3	25.0	36.3	48.2	76.3	83.4	89.6	94.2
	0.85	17.5	21.4	29.9	41.9	43.1	49.4	55.7	68.2
	0.35	11.2	14.3	21.6	34.3	28.9	36.2	40.3	53.8
1.0	9^9	51.9	55.7	63.8	72.9	84.7	92.3	99.4	99.9
	0.85	40.9	46.4	54.9	65.7	60.1	67.3	75.7	87.1
	0.35	20.1	24.3	34.5	48.8	39.6	44.6	54.7	69.7

In the second example we investigate the problem of testing for a parametric form of the scale function. In particular we consider the hypothesis

$$H_0 : \sigma(X) = \exp(\beta_1 + \beta_2 \log X) \tag{4.1}$$

and the regression model

$$Y_i = X_i + \exp(0.5 + \log X_i) \xi_i, \quad i = 1, \dots, n,$$

where X_1, \dots, X_n are i.i.d. $\sim \mathcal{U}[0, 1]$ and ξ_1, \dots, ξ_n are i.i.d. $\sim \mathcal{N}[0, 1]$. The censoring times are generated by the model

$$C_i = X_i + q + \exp(0.5 + \log X_i) \eta_i, \quad i = 1, \dots, n,$$

where q determines the amount of censoring and η_1, \dots, η_n are again $\sim \mathcal{N}[0, 1]$. The simulated rejection probabilities based on 1000 simulation runs are shown in Table 2 (the number of bootstrap replications is again $B = 199$ and the bandwidths are chosen as in the previous example). For the constant q we considered the cases $q = 9^9$, $q = 0.85$ and $q = 0.35$ which corresponds in the present context to an amount of 0%, 26% and 15% censoring under the null hypothesis. We observe a reasonable approximation of the nominal level, which is slightly worse compared to the hypothesis of homoscedasticity.

Table 2 *Simulated rejection probabilities of the bootstrap test for the parametric hypothesis (4.1) under the null hypothesis.*

n	50				100			
q	2.5%	5%	10%	20%	2.5%	5%	10%	20%
9^9	4.3	6.9	12.5	23.1	3.8	5.6	10.3	21.7
0.85	5.1	6.8	11.9	22.2	3.2	5.8	9.5	16.6
0.35	4.9	6.4	11.0	18.5	3.5	6.4	10.9	18.6

5 Data analysis

We are here interested in the (nonlinear) relationship between fatigue life of metal, ceramic or composite materials (which is considered as a survival time) and applied stress. This important input to design-for-reliability processes is motivated by the need to develop and present quantitative fatigue-life information used in the design of jet engines. Indeed, according to the air speed that enters an aircraft engine, the fan, the compressor and the turbine rotate at different speeds and therefore are submitted to different stresses. Moreover, fatigue life may be censored since failures may result from impurities or vacuums in the studied materials, or no failure may occur at all due to time constraints of the experiments.

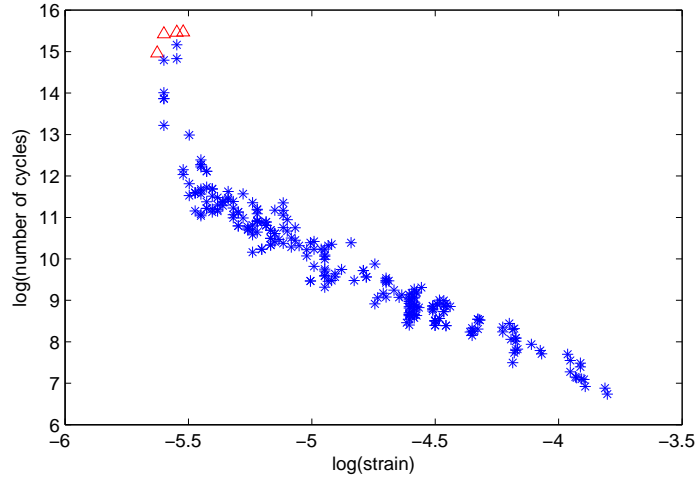


Figure 1: *Fatigue life data.* Scatter plot of the logarithms of fatigue life versus the logarithms of strain for specimens of a nickel-base superalloy. Uncensored data points are given by * and censored observations by Δ .

From a long time, an important question in fatigue analysis is to know whether the variability of fatigue life given the stress (or the strain) is constant for any stress (or strain). Several authors addressed this problem, among others, Nelson (1984) and Pascual and Meeker (1997,1999) who studied the number of cycles before failure of nickel-base superalloys as functions of the strain or the pseudostress (Young's modulus times strain). By example, Pascual and Meeker (1997) considered model (1.1) with the following form for the conditional standard deviation of the logarithm of the number of cycles before failure:

$$\sigma(X) = \exp(\beta_1 + \beta_2 \log X). \quad (5.1)$$

However, those authors assumed parametric forms for both $m(X)$ and the error distribution.

We present, in this section, a data set of 246 specimens of a nickel-base superalloy given by Shen (1994) and studied by Pascual and Meeker (1999). For these data, we consider model (1.1) where Y is the logarithm of the number of cycles before failure and X is the logarithm of the resulting strain (see Figure 1). Pascual and Meeker (1999) only use the 115 observations for which strain is below .007. The reason is that their completely parametric model doesn't fit the whole data set. As consequence, beyond robustness questions, there are obvious reasons to study $\sigma(X)$ independently of any parametric form.

In order to provide answers for the above questions, we display in the left part of Figure 2 the nonparametric estimates of the scale function $\sigma^2(x)$. Next we illustrate the

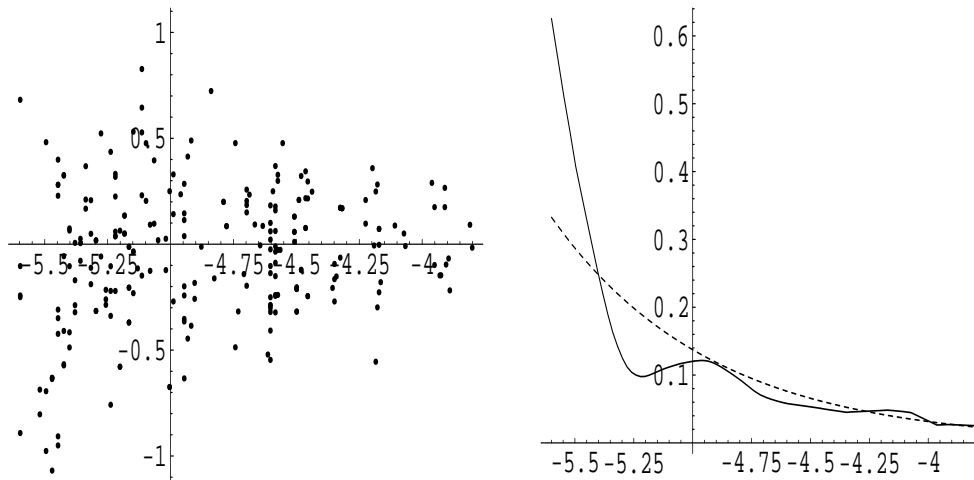


Figure 2: *Left panel: residuals of a nonparametric fit to the fatigue life data. Right panel: parametric and nonparametric estimates of the conditional scale function. Solid line: nonparametric estimate; dashed line: parametric estimate obtained under the hypothesis H_{02} .*

new test and consider the hypotheses $H_{0j} : \sigma(\cdot) \in \mathcal{M}_j, j = 1, 2$, for the classes of functions

$$\mathcal{M}_1 = \{\vartheta : \vartheta \in \Theta_1\},$$

(test for homoscedasticity) and

$$\mathcal{M}_2 = \{\exp(\beta_1 + \beta_2 \log X) : (\beta_1, \beta_2) \in \Theta_2\}$$

(test for the hypothesis (5.1)), where $\Theta_1 \subset \mathbb{R}$ and $\Theta_2 \subset \mathbb{R}^2$. For the problem of testing for a constant scale the new bootstrap test yields a p -value of 0.000 (using 499 bootstrap replications). This hypothesis is clearly rejected and reflects the picture of the residuals in Figure 1. For the hypothesis (5.1) the p -value of the test is slightly larger, i.e. 0.018 and the hypothesis cannot be rejected at the 1%-level. However, in the right part of Figure 2 we show the nonparametric estimate of the scale function (solid line) σ^2 and the corresponding parametric estimate (dashed line). This figure indicates that the function $\exp(\beta_1 + \beta_2 \log x)$ may not describe the variance structure adequately.

Acknowledgements. This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation (DFG).

Appendix

The following notations are needed in the statement of the asymptotic results given Section 3.

$$\begin{aligned}\xi(z, \delta, y|x) &= (1 - F(y|x)) \left\{ - \int_{-\infty}^{y \wedge z} \frac{dH_1(s|x)}{(1 - H(s|x))^2} + \frac{I(z \leq y, \delta = 1)}{1 - H(z|x)} \right\}, \\ \eta(z, \delta|x) &= \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x)) dv \sigma^0(x)^{-1}, \\ \zeta(z, \delta|x) &= \int_{-\infty}^{+\infty} \xi(z, \delta, v|x) J(F(v|x)) \frac{v - m^0(x)}{\sigma^0(x)} dv \sigma^0(x)^{-1}, \\ \gamma_0(y|x) &= \int_{-\infty}^y \frac{sh_\varepsilon^0(s|x)}{(1 - H_\varepsilon^0(s))^2} dH_{\varepsilon 1}^0(s) + \int_{-\infty}^y \frac{d(sh_{\varepsilon 1}^0(s|x))}{1 - H_\varepsilon^0(s)}, \\ \Omega &= E\left[\left(\frac{\partial \sigma_{\tilde{\theta}_0}(X)}{\partial \vartheta}\right)\left(\frac{\partial \sigma_{\tilde{\theta}_0}(X)}{\partial \vartheta'}\right)\right], \\ \rho(x, z, \delta, \tilde{\theta}_0) &= -\Omega^{-1} \sigma^0(x) \zeta(z, \delta|x) \frac{\partial \sigma_{\tilde{\theta}_0}(x)}{\partial \vartheta}, \\ \chi_{\tilde{\theta}_0}(x, z, \delta, y) &= (1 - F_\varepsilon^0(y)) [\gamma_0(y|x) \zeta(z, \delta|x) \\ &\quad + \int (\sigma^0(u))^{-1} \rho'(x, z, \delta, \tilde{\theta}_0) \frac{\partial \sigma_{\tilde{\theta}_0}(u)}{\partial \vartheta} \gamma_0(y|u) dF_X(u)].\end{aligned}$$

Let \tilde{T}_x be any value less than the upper bound of the support of $H(\cdot|x)$ such that $\inf_{x \in R_X} (1 - H(\tilde{T}_x|x)) > 0$. For a (sub)distribution function $L(y|x)$ we will use the notations $l(y|x) = L'(y|x) = (\partial/\partial y)L(y|x)$, $\dot{L}(y|x) = (\partial/\partial x)L(y|x)$ and similar notations will be used for higher order derivatives.

The assumptions needed for the asymptotic results are listed below.

(A1)(i) $na_n^4 \rightarrow 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \rightarrow \infty$ for some $\delta < 1/2$.

(ii) R_X is a compact interval of length L_X .

(iii) K is a density with compact support, $\int uK(u)du = 0$ and K is twice continuously differentiable.

(iv) The matrix Ω is non-singular.

(A2)(i) There exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_1 \leq \inf_x F(\tilde{T}_x|x)$, $s_0 \leq \inf\{s \in [0, 1]; J(s) \neq 0\}$, $s_1 \geq \sup\{s \in [0, 1]; J(s) \neq 0\}$ and $\inf_{x \in R_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$.

(ii) J is twice continuously differentiable, $\int_0^1 J(s)ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

(A3)(i) F_X is three times continuously differentiable and $\inf_{x \in R_X} f_X(x) > 0$.

(ii) m^0 is twice continuously differentiable.

(A4)(i) $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded, uniformly in $x \in R_X$, $z < \tilde{T}_x$ and δ .

(A5) For $L(y|x) = H(y|x), H_1(y|x), H_\varepsilon^0(y|x)$ or $H_{\varepsilon 1}^0(y|x) : L'(y|x)$ is continuous in (x, y) and $\sup_{x,y} |y^2 L'(y|x)| < \infty$, the same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order three.

(A6) For the density $f_{X|Z,\Delta}(x|z, \delta)$ of X given (Z, Δ) , $\sup_{x,z} |f_{X|Z,\Delta}(x|z, \delta)| < \infty$, $\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$, $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$ ($\delta = 0, 1$).

(A7) Θ is compact and $\tilde{\theta}_0$ is an interior point of Θ . All partial derivatives of $\sigma_\vartheta(x)$ with respect to the components of ϑ and x up to order three exist and are continuous in (x, ϑ) for all x and ϑ . Moreover, $\inf_{x \in R_X} \sigma^0(x) > 0$.

(A8) For all $\varepsilon > 0$, $\inf_{\|\vartheta - \tilde{\theta}_0\| > \varepsilon} E[(\sigma^0(X) - \sigma_\vartheta(X))^2] > 0$.

Lemma A.1 Assume a_n satisfies $na_n^5(\log a_n^{-1})^{-1} = O(1)$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \rightarrow \infty$ for some δ . Assume also (A1) (ii, iii), (A2) (i), J is continuous, $\int_0^1 J(s)ds = 1$, $J(s) \geq 0$ for all $0 \leq s \leq 1$, F_X is twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$, $H(y|x)$ and $H_1(y|x)$ satisfy (A5), Θ is compact and $\tilde{\theta}_0$ is an interior point of Θ , $\sigma_\vartheta(x)$ is continuous in (x, ϑ) for all x and ϑ , $\inf_{x \in R_X} \sigma^0(x) > 0$ and for all $\varepsilon > 0$, $\inf_{\|\vartheta - \tilde{\theta}_0\| > \varepsilon} E[(\sigma^0(X) - \sigma_\vartheta(X))^2] > 0$. Then under the null hypothesis H'_0 ,

$$\vartheta_n - \theta_0 \rightarrow_P 0.$$

Proof. Define $S_0(\vartheta) = E[(\sigma^0(X) - \sigma_\vartheta(X))^2]$. It follows from Theorem 5.7 in van der Vaart (1998, p. 45) that it suffices to show that

$$\sup_{\vartheta} |S_n(\vartheta) - S_0(\vartheta)| \rightarrow_P 0,$$

where $S_n(\vartheta) = (1/n) \sum_{i=1}^n (\hat{\sigma}^0(X_i) - \sigma_\vartheta(X_i))^2$. Using Proposition 4.5 of Van Keilegom and Akritas (1999) (hereafter abbreviated by VKA) enables to write

$$\sup_{\vartheta} |S_n(\vartheta) - S_0(\vartheta)| \leq \sup_{\vartheta} |\tilde{S}_n(\vartheta) - S_0(\vartheta)| + o_P(1),$$

where $\tilde{S}_n(\vartheta) = (1/n) \sum_{i=1}^n (\sigma^0(X_i) - \sigma_\vartheta(X_i))^2$. Next, the result follows from Theorem 2 in Jennrich (1969).

Lemma A.2 Assume (A1)-(A3), $\zeta(z, \delta|x)$ satisfies (A4), $H(y|x)$ and $H_1(y|x)$ satisfy (A5) and (A6)-(A8). Then under the null hypothesis H'_0 ,

$$\vartheta_n - \theta_0 = -\Omega^{-1}n^{-1} \sum_{i=1}^n \sigma^0(X_i) \zeta(Z_i, \Delta_i|X_i) \frac{\partial \sigma_{\theta_0}(X_i)}{\partial \vartheta} + o_P(n^{-1/2}).$$

Proof. For some ϑ_{1n} between ϑ_n and θ_0 ,

$$\vartheta_n - \theta_0 = -\left\{ \frac{\partial^2 S_n(\vartheta_{1n})}{\partial \vartheta \partial \vartheta'} \right\}^{-1} \frac{\partial S_n(\theta_0)}{\partial \vartheta},$$

where

$$\begin{aligned} \frac{\partial S_n(\theta_0)}{\partial \vartheta} &= -2n^{-1} \sum_{i=1}^n (\hat{\sigma}^0(X_i) - \sigma_{\theta_0}(X_i)) \frac{\partial \sigma_{\theta_0}(X_i)}{\partial \vartheta} \\ &= -2n^{-1} \sum_{i=1}^n (\hat{\sigma}^0(X_i) - \sigma^0(X_i)) \frac{\partial \sigma_{\theta_0}(X_i)}{\partial \vartheta}. \end{aligned}$$

We have by Proposition 4.9 of VKA (1999)

$$\begin{aligned} \frac{\partial S_n(\theta_0)}{\partial \vartheta} &= 2n^{-2} a_n^{-1} \sum_{i=1}^n \sum_{j=1}^n K\left(\frac{X_i - X_j}{a_n}\right) f_X^{-1}(X_i) \sigma^0(X_i) \zeta(Z_j, \Delta_j|X_i) \frac{\partial \sigma_{\theta_0}(X_i)}{\partial \vartheta} + o_P(n^{-1/2}) \\ &= 2n^{-1} \sum_{j=1}^n \sigma^0(X_j) \zeta(Z_j, \Delta_j|X_j) \frac{\partial \sigma_{\theta_0}(X_j)}{\partial \vartheta} + o_P(n^{-1/2}), \end{aligned}$$

using arguments similar to those used in expressions (A.5) to (A.7) of Heuchenne and Van Keilegom (2007). Next,

$$\begin{aligned} &\frac{\partial^2 S_n(\vartheta_{1n})}{\partial \vartheta \partial \vartheta'} \\ &= -2n^{-1} \left\{ \sum_{i=1}^n (\hat{\sigma}^0(X_i) - \sigma_{\vartheta_{1n}}(X_i)) \frac{\partial^2 \sigma_{\vartheta_{1n}}(X_i)}{\partial \vartheta \partial \vartheta'} - \sum_{i=1}^n \left(\frac{\partial \sigma_{\vartheta_{1n}}(X_i)}{\partial \vartheta} \right) \left(\frac{\partial \sigma_{\vartheta_{1n}}(X_i)}{\partial \vartheta'} \right) \right\} \\ &= 2\Omega + o_P(1), \end{aligned}$$

for which Lemma A.1, Proposition 4.5 of VKA (1999) and assumption (A7) are used. This finishes the proof.

Lemma A.3 If the assumptions of Lemma A.2 are satisfied, then under the null hypothesis H'_0 , $n^{1/2}(\vartheta_n - \theta_0) \xrightarrow{d} N(0, \Sigma)$, where

$$\Sigma = \Omega^{-1} E\left[\sigma^{02}(X) \zeta^2(Z, \Delta|X) \frac{\partial \sigma_{\theta_0}(X)}{\partial \vartheta} \frac{\partial \sigma_{\theta_0}(X)}{\partial \vartheta'} \right] \Omega^{-1}.$$

Proof. The proof follows readily from Lemma A.2.

Proof of Theorem 3.1. The objective of the proof is to provide asymptotic representations for both differences $\hat{F}_{\varepsilon_0}(y) - F_{\varepsilon_0}(y)$ and $\hat{F}_{\varepsilon}^0(y) - F_{\varepsilon}^0(y)$. Indeed, under H'_0 , $F_{\varepsilon_0}(y) = F_{\varepsilon}^0(y)$, for $-\infty < y \leq T$. Following the lines of Theorem 3.1 of VKA (1999), we obtain

$$\begin{aligned} & \hat{F}_{\varepsilon_0}(y) - F_{\varepsilon_0}(y) \\ &= (1 - F_{\varepsilon_0}(y)) \left[\int_{-\infty}^y \frac{\hat{H}_{\varepsilon_0}(s) - H_{\varepsilon_0}(s)}{(1 - H_{\varepsilon_0}(s))^2} dH_{\varepsilon_0}(s) + \int_{-\infty}^y \frac{d(\hat{H}_{\varepsilon_0}(s) - H_{\varepsilon_0}(s))}{1 - H_{\varepsilon_0}(s)} \right] \\ &+ R_{n0}(y) \end{aligned} \tag{A.1}$$

and

$$\begin{aligned} & \hat{F}_{\varepsilon}^0(y) - F_{\varepsilon}^0(y) \\ &= (1 - F_{\varepsilon}^0(y)) \left[\int_{-\infty}^y \frac{\hat{H}_{\varepsilon}^0(s) - H_{\varepsilon}^0(s)}{(1 - H_{\varepsilon}^0(s))^2} dH_{\varepsilon}^0(s) + \int_{-\infty}^y \frac{d(\hat{H}_{\varepsilon}^0(s) - H_{\varepsilon}^0(s))}{1 - H_{\varepsilon}^0(s)} \right] \\ &+ R_n^0(y), \end{aligned} \tag{A.2}$$

where $\sup\{|R_{n0}(y)|; -\infty < y \leq T\} = o_P(n^{-1/2})$, $\sup\{|R_n^0(y)|; -\infty < y \leq T\} = o_P(n^{-1/2})$, $\hat{H}_{\varepsilon_0}(y)$, $\hat{H}_{\varepsilon_0}(y)$, $\hat{H}_{\varepsilon}^0(y)$, $\hat{H}_{\varepsilon_1}^0(y)$ denote the estimators

$$(1/n) \sum_{i=1}^n I(\hat{\mathcal{E}}_i \leq y)$$

of the (sub)distributions $H_{\varepsilon_0}(y)$, $H_{\varepsilon_0}(y)$, $H_{\varepsilon}^0(y)$, $H_{\varepsilon_1}^0(y)$ for $\hat{\mathcal{E}}_i = \hat{E}_{i,0}^0$, $\hat{E}_{i,0}^0$ for which $\Delta_i = 1$, \hat{E}_i^0 and \hat{E}_i^0 for which $\Delta_i = 1$, $i = 1, \dots, n$.

Now, from the proof of Proposition A.2 in VKA (1999), we have that

$$\begin{aligned} \hat{H}_{\varepsilon_0}(y) - H_{\varepsilon_0}(y) &= \frac{1}{n} \sum_{i=1}^n (I(E_{i,0}^0 \leq y) - H_{\varepsilon_0}(y)) \\ &+ \int h_{\varepsilon_0}(y|x) \frac{\hat{m}^0(x) - m^0(x)}{\sigma^0(x)} dF_X(x) \\ &+ \int y h_{\varepsilon_0}(y|x) \frac{\sigma_{\partial_n}(x) - \sigma^0(x)}{\sigma^0(x)} dF_X(x) + o_P(n^{-1/2}), \end{aligned} \tag{A.3}$$

uniformly in $-\infty < y \leq T$. The last term (lower order terms of a Taylor development) is $o_P(n^{-1/2})$ because of assumption (A7), Proposition 4.5 of VKA (1999), Lemma A.3 and the fact that $\sup_{x,y} |y^2 h'_{\varepsilon_0}(y|x)| < \infty$. In the same way,

$$\hat{H}_{\varepsilon_0}(y) - H_{\varepsilon_0}(y) = \frac{1}{n} \sum_{i=1}^n (I(E_{i,0}^0 \leq y, \Delta_i = 1) - H_{\varepsilon_0}(y))$$

$$\begin{aligned}
& + \int h_{\varepsilon_0 1}(y|x) \frac{\hat{m}^0(x) - m^0(x)}{\sigma^0(x)} dF_X(x) \\
& + \int y h_{\varepsilon_0 1}(y|x) \frac{\sigma_{\vartheta_n}(x) - \sigma^0(x)}{\sigma^0(x)} dF_X(x) + o_P(n^{-1/2}), \quad (\text{A.4})
\end{aligned}$$

uniformly in $-\infty < y \leq T$, and similarly for both $\hat{H}_\varepsilon^0(y) - H_\varepsilon^0(y)$ and $\hat{H}_{\varepsilon_1}^0(y) - H_{\varepsilon_1}^0(y)$, where $E_{i,0}^0$ (resp. $H_{\varepsilon_0}(y)$, $h_{\varepsilon_0}(y|x)$, $H_{\varepsilon_0 1}(y)$ and $h_{\varepsilon_0 1}(y|x)$) is replaced by E_i^0 (resp. $H_\varepsilon^0(y)$, $h_\varepsilon^0(y|x)$, $H_{\varepsilon_1}^0(y)$ and $h_{\varepsilon_1}^0(y|x)$), $i = 1, \dots, n$.

The results (A.1), (A.3) and (A.4) are actually true if Lemma A.1 of VKA (1999) can be used with functions $d_2(x) \in \Sigma_\Theta(R_X) = \{x \rightarrow \frac{\sigma_\vartheta(x)}{\sigma^0(x)} \text{ with } \inf_x \{\frac{\sigma_\vartheta(x)}{\sigma^0(x)}\} \geq \frac{1}{2}, \vartheta \in \Theta\}$ instead of $d_2(x) \in \tilde{C}_2^{1+\delta}(R_X)$ used in this Lemma. It is clear that $P(\frac{\sigma_{\vartheta_n}(x)}{\sigma^0(x)} \in \Sigma_\Theta(R_X)) \rightarrow 1$ as $n \rightarrow \infty$. Next, the bracketing number $N_{[]}(\lambda^2, \Sigma_\Theta(R_X), L_2(P)) = O(\lambda^{-2D})$ for any $\lambda > 0$, due to the compactness of $\Theta \in \mathbb{R}^D$. Since this bracketing number is smaller than for the class $\tilde{C}_2^{1+\delta}(R_X)$, $\tilde{C}_2^{1+\delta}(R_X)$ can be replaced by $\Sigma_\Theta(R_X)$ in Lemma A.1 of VKA (1999) and consequently, (A.1), (A.3) and (A.4) are true.

Next, we treat the right hand side of (A.3). (A.4) is treated similarly. Using Proposition 4.8 of VKA (1999) and a Taylor development, we obtain for the second term of the right hand side of (A.3)

$$\int h_{\varepsilon_0}(y|x) \frac{\hat{m}^0(x) - m^0(x)}{\sigma^0(x)} dF_X(x) = -n^{-1} \sum_{i=1}^n h_{\varepsilon_0}(y|X_i) \eta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}), \quad (\text{A.5})$$

uniformly in $-\infty < y \leq T$. By Lemma A.2 and a Taylor development, the third term on the right hand side of (A.3) becomes

$$n^{-1} \sum_{i=1}^n \int y h_{\varepsilon_0}(y|x) (\sigma^0(x))^{-1} \sum_{d=1}^D \rho_d(X_i, Z_i, \Delta_i, \theta_0) \frac{\partial \sigma_{\theta_0}(x)}{\partial \vartheta_d} dF_X(x) + o_P(n^{-1/2}), \quad (\text{A.6})$$

uniformly in $-\infty < y \leq T$, where $\rho_d(X, Z, \Delta, \theta_0)$ is the d^{th} component of the vector $\rho(X, Z, \Delta, \theta_0)$, $d = 1, \dots, D$. Substituting (A.5) and (A.6) in (A.3) leads to

$$\begin{aligned}
\hat{H}_{\varepsilon_0}(y) - H_{\varepsilon_0}(y) & = \frac{1}{n} \sum_{i=1}^n (I(E_{i,0}^0 \leq y) - H_{\varepsilon_0}(y)) \\
& - n^{-1} \sum_{i=1}^n h_{\varepsilon_0}(y|X_i) \eta(Z_i, \Delta_i|X_i) \\
& + n^{-1} \sum_{i=1}^n \int \left\{ y h_{\varepsilon_0}(y|x) (\sigma^0(x))^{-1} \right. \\
& \quad \left. \times \sum_{d=1}^D \rho_d(X_i, Z_i, \Delta_i, \theta_0) \frac{\partial \sigma_{\theta_0}(x)}{\partial \vartheta_d} \right\} dF_X(x) \\
& + o_P(n^{-1/2}), \quad (\text{A.7})
\end{aligned}$$

uniformly in $-\infty < y \leq T$. In the same way, (A.4) has the same structure with $I(E_0^0 \leq y)$ (resp. $H_{\varepsilon_0}(y)$ and $h_{\varepsilon_0}(y|x)$) replaced by $I(E_0^0 \leq y, \Delta = 1)$ (resp. $H_{\varepsilon_01}(y)$ and $h_{\varepsilon_01}(y|x)$).

Finally, Proposition A.2 of VKA (1999) ensures that

$$\begin{aligned} \hat{H}_\varepsilon^0(y) - H_\varepsilon^0(y) &= \frac{1}{n} \sum_{i=1}^n (I(E_i^0 \leq y) - H_\varepsilon^0(y)) \\ &\quad - n^{-1} \sum_{i=1}^n h_\varepsilon^0(y|X_i) \eta(Z_i, \Delta_i|X_i) \\ &\quad - n^{-1} \sum_{i=1}^n y h_\varepsilon^0(y|X_i) \zeta(Z_i, \Delta_i|X_i) + o_P(n^{-1/2}), \end{aligned} \quad (\text{A.8})$$

uniformly in $-\infty < y \leq T$, and similarly for $\hat{H}_{\varepsilon_1}^0(y) - H_{\varepsilon_1}^0(y)$. Therefore introducing (A.7), (A.8) and their counterparts for $\hat{H}_{\varepsilon_01}(y) - H_{\varepsilon_01}(y)$ and $\hat{H}_{\varepsilon_1}^0(y) - H_{\varepsilon_1}^0(y)$ in (A.1) and (A.2) leads to the asymptotic representation proposed in Theorem 3.1.

Proof of Theorem 3.2. We will make use of Theorem 2.5.6 in van der Vaart and Wellner (1996), i.e. we will show that

$$\int_0^\infty \sqrt{\log N_{[]}(\lambda, \mathcal{F}, L_2(P))} d\lambda < \infty, \quad (\text{A.9})$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of (X, Z, Δ) , $L_2(P)$ is the L_2 -norm and $\mathcal{F} = \{\chi_{\theta_0}(X, Z, \Delta, y); -\infty < y \leq T\}$. Proving (A.9) implies that \mathcal{F} is a Donsker class and hence the weak convergence of the given process is ensured by pages 81-82 of van der Vaart and Wellner's book. First, the functions $x \rightarrow (1 - F_\varepsilon^0(y))\gamma_0(y|x)$ are bounded uniformly in y as well as their first derivatives since $\sup_{x,y} |y \dot{h}_\varepsilon^0(y|x)| < \infty$ and $\sup_{x,y} |y \dot{h}_{\varepsilon_1}^0(y|x)| < \infty$. By Corollary 2.7.2 of the aforementioned book, their bracketing number is $O(\exp(K\lambda^{-1}))$. Since $\zeta(z, \delta|x)$ is uniformly bounded, the bracketing number of the first term of $\chi_{\theta_0}(x, z, \delta, y)$ is $O(\exp(K\lambda^{-1}))$. Next, the second term of $\chi_{\theta_0}(x, z, \delta, y)$ is divided into D terms corresponding to each term of the scalar product $\rho'(x, z, \delta, \theta_0) \frac{\partial \sigma_{\theta_0}(x)}{\partial \vartheta}$. Each term can therefore be written

$$\rho_d(x, z, \delta, \theta_0) (1 - F_\varepsilon^0(y)) \left[\int (\sigma^0(u))^{-1} \frac{\partial \sigma_{\theta_0}(u)}{\partial \vartheta_d} \gamma_0(y|u) dF_X(u) \right],$$

for $d = 1, \dots, D$, which is immediately treated since it is factorized in a (uniformly bounded) function independent of y and a uniformly bounded function only depending on y . This concludes the proof since the integration in (A.9) can be restricted to the interval $[0, 2M]$, since $|\chi_{\theta_0}(x, z, \delta, y)| \leq M$ (for a specific choice of M), for all x, z, δ and y (for $\lambda > 2M$, we take $N_{[]}(\lambda, \mathcal{F}, L_2(P)) = 1$).

Proof of Corollary 3.3. The convergence of T_{KS} follows directly from the weak convergence of the process $\hat{W}(y)$ and the continuous mapping Theorem. For T_{CM} , write

$$\begin{aligned} & \int_{-\infty}^T \hat{W}^2(y) d\hat{F}_\varepsilon^0(y) - \int_{-\infty}^T W^2(y) dF_\varepsilon^0(y) \\ & \leq \int_{-\infty}^T (\hat{W}^2(y) - W^2(y)) d\hat{F}_\varepsilon^0(y) + \int_{-\infty}^T W^2(y) d(\hat{F}_\varepsilon^0(y) - F_\varepsilon^0(y)). \end{aligned}$$

For the first term on the right hand side of the above inequality, we apply the Skorohod construction (see Serfling, 1980) to the process $\hat{W}(y)$ such that $\sup_{-\infty < y \leq T} |\hat{W}(y) - W(y)| \rightarrow 0$, *a.s.* The second term is jointly treated by the Skorohod construction applied to the process $n^{1/2}(\hat{F}_\varepsilon^0(y) - F_\varepsilon^0(y))$ and the Helly-Bray Theorem (see p. 97 in Rao, 1965) applied to each of the trajectories of $W(y)$. Finally, we use Corollary 3.2 of VKA (1999) to treat the difference

$$\frac{1}{\hat{F}_\varepsilon^0(T)} \int_{-\infty}^T W^2(y) dF_\varepsilon^0(y) - \frac{1}{F_\varepsilon^0(T)} \int_{-\infty}^T W^2(y) dF_\varepsilon^0(y).$$

References

- Atkinson, A. C. (1985). *Plots, transformations and regression : an introduction to graphical methods of diagnostic*. Oxford : Clarendon Pr.
- Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.
- Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of Statistical Planning and Inference*, **103**, 311-330
- Dette, H. and Hetzler, B. (2009). Khmaldaze transformation of integrated variance processes with applications to goodness-of-fit testing. *Mathematical Methods of Statistics*, **18(2)**, 97-116.
- Dette, H. and Hetzler, B. (2009). A simple test for the parametric form of the variance function in nonparametric regression. To appear in *Annals of the Institute of Statistical Mathematics*, (DOI 10.1007/s10463-008-0169-1).
- Dette, H. , Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *J. R. Statist. Soc., Ser. B*, **60**, 693-708.
- Dette, H., Pardo Fernández, J. C. and Van Keilegom, I. (2007). Goodness-of-fit tests for multiplicative models with dependent data. To appear in *Scandinavian Journal of Statistics*.

- Dette, H. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *J. Royal Statist. Soc. - Series B.* **69**, 903–917.
- Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.*, **76**, 312–319.
- Feller, W. (1966). *An introduction to probability theory and its applications. Vol II.* Wiley Series in Probability and Mathematical Statistics. New York.
- González Manteiga, W., Heuchenne, C. and Sánchez Sellero, C. (2009). Goodness-of-fit tests for censored regression based on artificial data points. *Submitted paper.*
- Heuchenne, C. and Van Keilegom, I. (2007). Polynomial regression with censored data based on preliminary nonparametric estimation. *Ann. Instit. Statist. Math.*, **59**, 273–298.
- Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.*, **40**, 633–643.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.
- Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, **2**, 164–168.
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.
- Nelson, W. (1984). Fitting of fatigue curves with nonconstant standard deviation to data with runouts. *J. Testing Eval.*, **12**, 69–77.
- Pardo Fernández, J. C., Van Keilegom, I. and González Manteiga, W. (2007). Goodness-of-fit tests for parametric models in censored regression. *Canad. J. Statist.*, **35**, 249–264.
- Pascual, F. G. and Meeker, W. Q. (1997). Analysis of fatigue data with runouts based on a model with nonconstant standard deviation and a fatigue limit parameter. *J. Testing Eval.*, **25**, 292–301.
- Pascual, F. G. and Meeker, W. Q. (1999). Estimating Fatigue Curves With the Random Fatigue-Limit Model. *Technometrics*, **4**, 277–290.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.
- Shen, C. L., (1994). *Statistical Analysis of Fatigue Data.* Unpublished Ph.D. dissertation, University of Arizona, Dept. of Aerospace and Mechanical Engineering.
- van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer, New-York.
- Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored

regression models. *Ann. Statist.*, **27**, 1745–1784.

Van Keilegom, I., González Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, **17**, 401–415.

You, J., Chen, G. (2005). "Testing heteroscedasticity in partially linear regression models", *Statist. Probab. Lett.*, **73**, 61–70.